

PRICE'S INDEX THROUGH OF WEB SCRAPING

ÍNDICE DE PRECIOS A TRAVÉS DE WEB SCRAPING

Paulina Pegueroles*

Rubén Guerrero**

Amaru Fernández***

Diana López****

RESUMEN:

Propósito: este estudio presenta la elaboración de un Índice de Precios Inmobiliarios para la Región Metropolitana en el período que comprende desde enero de 2017 al 14 de agosto de 2019. El índice, que mide el dinamismo del sector, se construyó procesando 750.000 observaciones obtenidas desde el sitio web TocToc.com. Este sitio es una fuente secundaria de información que proporciona variables deterministas para construir una base de datos formal, confiable y representativa.

Diseño / Metodología / Enfoque: para generar el Índice de Precios Inmobiliarios en la Región Metropolitana, adoptamos una metodología basada en el índice sintético, en particular el Modelo de Precios Hedónicos (HPM), que se explica en términos de características (heterogéneas).

El desafío será agregar más observaciones a la muestra recopilada del sitio web TocToc y otros relacionados.

Hallazgos: la metodología de Laspeyres mostró los mejores resultados, pero una volatilidad marginal en comparación con las otras técnicas. Además, los resultados presentaron una tendencia a la baja en los precios finales de los hogares

en el segundo trimestre de 2019 debido al bajo número de transacciones y al estancamiento del sector de la construcción, mostrando resultados similares a los proporcionados por el Banco Central de Chile.

Originalidad / Valor: el uso del web scraping como herramienta para obtener los datos en tiempo real, lo que permite analizar los precios de los bienes inmuebles en este momento, es una ventaja, ya que otros índices tienen un retraso de seis meses o más.

PALABRAS CLAVES: índice de precios, web scraping, construcción, Modelo Hedónico, Características Heterogéneas, Índice de Laspeyres

*Universidad Tecnológica Metropolitana, UTEM. Santiago, Chile. Paulina.peguerolese@utem.cl

** Universidad Tecnológica Metropolitana, UTEM. Santiago, Chile. Ruben.guerrero@utem.cl

*** Universidad Tecnológica Metropolitana, UTEM. Santiago, Chile. Amarud@gmail.com

**** Universidad Tecnológica Metropolitana, UTEM. Santiago, Chile. Diana.lopez@utem.com

ABSTRACT:

Purpose: This study presents a design for a Real Estate Price Index for Región Metropolitana in the period January 2017 to August 14 2019. The Index measures the dynamism of the sector and it was built by processing 750,000 observations obtained from the website TocToc.com. This site is a secondary source of information providing deterministic variables to build a formal, reliable and representative database.

Design/ Methodology/ Approach: In order to generate the Real Estate Price Index in Región Metropolitana, we adopted a methodology based on synthetic index, particularly the Hedonic Price Model (HPM), what is explained in terms of characteristics (heterogeneous).

The challenge will be to add more observations to the sample collected from the website TocToc and others related.

Findings: The Laspeyres methodology showed the highest results, but marginal volatility compared to the other techniques. In addition, the results presented a declining trend on household final prices in the second quarter of 2019 due to both the low number of transactions, and the stagnation

of the construction sector, showing similar results to the ones provided by Central Bank of Chile.

Originality/ Value: The use of web scraping as a tool to obtain the data in real time, allowing to analyze Real State prices at the moment, is an advantage, because other indexes have a six month or more time of delay.

KEY WORDS: Real Estate Price Index, Web scraping, Construction, Hedonic Model, characteristics heterogeneous, Laspeyres Index.

Article Type: Research paper

JEL Code: E31

Fecha de recepción: 20 de marzo de 2021

Fecha de aceptación: 27 de abril de 2021

INTRODUCTION AND RESEARCH PROBLEM

This study presents a design for a Real Estate Price Index for Región Metropolitana in the period January 2017 to August 14 2019. The Index measures the dynamism of the sector and it was built by processing 750,000 number of observations obtained from the website Toc Toc.com. This site is a secondary source of information providing deterministic variables to build a formal, reliable and representative database.

The Real Estate market is key for the growth of some countries¹ due to its role in the gross generation of fixed capital, consumption, and the financial system. Any variation in housing prices impacts the nations' added demand and its financial stability, as Idrovo & Lennon state:

En Chile y en la mayoría de los países desarrollados la inversión inmobiliaria cumple un rol fundamental en la actividad agregada. Tal es así que en la composición de Cuentas Nacionales del Banco Central se considera a la vivienda como un componente de la Formación Bruta de Capital Fijo, equivalente al 29% del valor corriente de la inversión en construcción (esto es un tercio de su valor real) y a 2,4% del PIB agregado (2011, p. 3).

Central Bank of Chile (2014) have developed a real estate price index, showing biased estimates because of the diversity of property, posing a need for a set of uniform objects to remove or minimize this bias. On the other hand, dwellings are characterized by the diversity of their characteristics, making them unique and incomparable goods. This aspect presents the main difficulty to elaborate a real estate price index.

The development of an index depends on applying different methodologies to decrease the bias, as Idrovo & Lennon (2011) state:

Estos sesgos sólo podrían eliminarse si se comparan en cada período los precios de mercado de exactamente los mismos inmuebles, lo cual es empíricamente imposible, por lo que surge la necesidad de utilizar métodos econométricos que permitan construir en forma teórica lo que no puede sostenerse en la práctica (2011, p. 4).

The problem stands, therefore, on how to analyse housing price variations considering the dwellings' attributes and minimizing the bias in the data, as found in previous research carried out by Banco Central de Chile (BCCh) and Cámara Chilena de la Construcción (CChC).

OBJECTIVES

The main objective of the research is to prove that a secondary source of information can be used to represent the situation in the sector timely and effectively² similarly to a primary one. It also aims at showing that features such as price, square meters and the location of the dwelling, are not the main deterministic variables for designing a real estate housing index.

The information provided by this real estate index could potentially be used by any agent in the economy, either for decision-making in the sector, investors or by individuals buying property. For the time being, the data will be available for generating timely statistics of housing price dynamism, so that the strength of the index can be confirmed through time.

1. An example is the 2007-2009 Sub-Prime Crisis in the US where Banks offered mortgages aiming at the geographic diversification of their investments, adding a greater spreading the risk of their portfolio. This fact generated a real estate bubble directly affecting the sector and other agents of economy.

2. It is a source of information that can explain the dynamism of the real estate sector with real, timely and current data.

THEORETICAL FRAMEWORK

The housing market, the real estate and the construction sector shape a market as a whole. However, there are some subtle differences among them. The construction industry produces dwellings of varying nature with different degrees of intervention. Housing, in turn, is judged to be a free market, self-regulated by suppliers and demanders in the sector. Quite the opposite, public works are embedded in non-residential areas and are entirely intervened.

In this line, Obaíd (2003) presents typical characteristics of buildings, particularly new and old, in the real estate sector:

[...] una de ellas y es evidente es el caso de las transacciones, a diferencia de otros mercados las transacciones son mucho menores, también los activos sólo pueden valorarse por referencias y se caracteriza por una baja capacidad de generar liquidez a corto plazo, esto por ser activos fijos y su capacidad de realización es lenta y difícil. Lo cual se debe a todo el trámite legal que se debe seguir y a los elevados precios de las operaciones, incluso las más pequeñas, si se comparan con otros mercados (Salazar & Díaz, 2014, p. 17).

In Chile, as well as other countries, dwellings are an essential component of family wealth, and the main loan guarantee in the financial system. For this reason, the variation of housing prices is directly related to the household consumption, affecting the financial situation of the country and the financing entities (Parrado, Cox & Fuenzalida, 2009).

As a result, the stability of the financial system is fundamental because it offers a high capacity to grant loans to real estate companies and partnerships. This fact was evident in 2007, when the construction sector represented 9% placement of the whole banking system. The real estate market is, therefore, relevant for the economic stability of the country, both in its role for household balance and for the financial system. It

has a positive impact generating employment in the construction sector, which was close to 10% of the total workforce in 2017, and it is a contribution to the nominal GDP, reaching 14% in this period (Ortúzar, 2018).

METHODOLOGY

In order to generate the real estate price index in Región Metropolitana, we adopted a methodology based on synthetic index, particularly the **Hedonic Price Model (HPM)**, due to the fact that the *repeat purchase rate* must be applied when there are price changes in dwellings sold more than once (Banco central, 20014, p.22). In addition, this method involves only old homes (Idrovo & Lennon, 2011), unlike the HPM model, where the price is explained in terms of characteristics (heterogeneous for each dwelling); i.e., the effect of every characteristic on the dwelling price is estimated by means of multivariate analysis (GLS³). Additionally, it does not distinguish between new and old homes.

This method can present disadvantages since it could generate biased estimates for two reasons: first, it avoids relevant variables directly affecting the dwelling price; second, there is an inadequate relation between specific characteristics of dwellings and its effects on the final price. Finally, it considers that dwelling peculiarities are invariable through time, which is mistaken.

The analysis is carried out through three models taking into account the following variables in their generation:

- UF (Unidad de Fomento) value
- Useful square metres (m²)
- Bathrooms
- Bedrooms
- Operation type (house/apartment; new/old)
- District groups in Region Metropolitana (RM)
- Advertisement date (month and year)

3. Generalised Least Square.

Additionally, the following interactive variables were considered:

1) Area interactive variable: the multiplication of district group and useful square metres

$$\text{Area}_n = \text{group}_n \cdot m_2 u$$

2) Where Area_n is the result of the multiplication of district groups of Región Metropolitana (group_n) and the useful square metres ($m_2 u$).

3) Type of dwelling interactive variable: the multiplication of the dummy variable house or apartment and m_2

$$C = m_2 u \text{ house/apartment}$$

C is the result of the multiplication of the dummy variable house or apartment (**house/apartment**) and the useful square metres (m_2).

The models are expressed as follows:

Equation 1: Multivariate Regression: (r1):

$$P_i = \beta_0 + \beta_1 \cdot m_2 u + \beta_2 \cdot \text{Bathrooms} + \beta_3 \cdot \text{Bedrooms} + \beta_4 \cdot \text{New/old} + \beta_5 \cdot \text{House/apartment} + \beta_6 \cdot \text{Group}_n + \beta_7 \cdot \text{Area}_n + \beta_8 \cdot C + \mu_i$$

Where the dwelling price (P_i) is explained by means of their different characteristics.

Equation 2: Multivariate Regression: (r2):

$$\text{Ln}P_i = \beta_0 + \beta_1 \cdot m_2 u + \beta_2 \cdot \text{Bathrooms} + \beta_3 \cdot \text{Bedrooms} + \beta_4 \cdot \text{New/old} + \beta_5 \cdot \text{House/apartment} + \beta_6 \cdot \text{Group}_n + \beta_7 \cdot \text{Area}_n + \beta_8 \cdot C + \mu_i$$

The second model or semi-log (r2) transforms the dependent variable into a logarithm.

Equation 3: Multivariate Regression: (r3):

$$\ln P_i = \beta_0 + \beta_1 \ln_{mt2u} + \beta_2 \cdot \text{Bathrooms} + \beta_3 \cdot \text{Bedrooms} + \beta_4 \cdot \text{New/old} + \beta_5 \cdot \text{House/apartment} + \beta_6 \cdot \text{Group}_n + \beta_7 \cdot \text{Area}_n + \beta_8 \cdot C + \mu_i$$

The third and last model log-log (r3) expresses the price and the useful square metres in a logarithm.

The Akaike Information Criterion (AIC) was used to select one of the three models, showing that the log-log (r3) explains data with the minimum number of parameters, in comparison to the linear model, which takes into account the non-linearity between the square metres and the dwelling price, as shown in Table 1.

The coefficients and standard deviations for every explanatory model variable can be observed in Table 2. In tables 3 and 4, it is possible to observe the single coefficients for both apartment and house index.

After the selection of the model, the price index was formulated based on three kinds of methodology: Laspeyres, Paasche and Fisher

- The Laspeyres Index⁴) refers to an arithmetic mean of simple price index, used as estimation for the value of transactions carried out in a base period; i.e., in every period t (Curiel Díaz, J., 1997).
- The Paasche Index⁵ (takes into account the purchase patterns of regular buyers, showing the taste and needs of consumers within the index (Webster A., 2000).
- The Fisher Index is the ideal price index because it reduces the shortcomings resulting from the other

4. The advantage of this index is that it requires data for only one period because its only varying element in time is price. However, it is disadvantageous in that it overestimates the increasing good prices, keeping the amount constant.

5. The weakness of this index is to underestimate increasing good prices.

two index, finding the square root in its product (Webster A., 2000).

In addition, the models have in common the same parameters and corresponding to the slope and intercept, respectively.

Equation 4: Price Index

$$I_{t_n}^l = \frac{E\left(\frac{p_{it}}{\bar{x}_{t_0}}, \vartheta_{t_n}, \theta_{t_n}\right)}{E\left(\frac{p_{it}}{\bar{x}_{t_0}}, \vartheta_{t_0}, \theta_{t_0}\right)}$$

$$(1) \quad I_{t_n}^p = \frac{E\left(\frac{p_{it}}{\bar{x}_{t_n}}, \vartheta_{t_n}, \theta_{t_n}\right)}{E\left(\frac{p_{it}}{\bar{x}_{t_n}}, \vartheta_{t_0}, \theta_{t_0}\right)}$$

$$I_{t_n}^f = (I_{t_n}^l I_{t_n}^p)^{1/2}$$

Source: Cámara Chilena de la Construcción, 2011.

RESULTS

The observations used for the Real Estate Index of the RM data were reduced from a total of 750,000 to 85,798, because the following dwellings were removed: all dwellings for rent, all those outside RM, and those that had no data in relation to region, district, useful square metres, bedrooms, bathrooms, UF value, and those observations generating data outliers. 16% of total data were imputed because they had data in the other variables and they lacked one characteristic. Table N°5 shows the percentage of the imputed variables.

The UF value is the dependent variable of the study. It was modified according to the value stated by Banco Central de Chile at the time the dwelling was adver-

tised. Then, all missing values was imputed⁶, under 150 UF and over 55,000 UF. Another variable used to establish maximum and minimum margins is the useful square metres, with a minimum of 21 m² and a maximum of 10,000 m². These margins were selected because some square metres in dwellings and the area they were located that in were not regulated, either because of mistyping or a wrong report on the dwellings' characteristics.

After the imputation, we can see that out of the 85,798 observations, 49% are houses and 51% apartments. 1% correspond to new housing and 99% to old ones. This shows that they are mostly old apartments.

To make a difference between new and old real estate, we created a dummy where 1 is new and 0 is old. The same was done with houses being 1 and apartments being 0.

Regarding the number of observations, 2019 is the year with the highest number of advertised real estate with a total of 43,759, which represents a 25% increase compared to 2018. In 2018 the number of advertisements had an increase 4 times higher than the previous year, probably due to the tendencies of e-commerce in Chile. According to Centro de Economía Digital CCS (2019) financial and real estate services had an increase of 146% in the digital market in 2018. 57 districts comprise the database. They are classified into seven groups according to their frequency rate and their geographical location. Table N°7 shows these frequencies.

From the above table, we can see that group 1 has the highest number of advertisements with real estate for sale with 30,839. Las Condes is the district with the highest number of observations with a total of 13,576, followed by Santiago with 9,206 data in group 4.

With the district groups it is possible to observe the relation between square metres and the UF value. See the Dispersion graph 1.

This graph displays that the data of group 1 lean towards the left, showing an important number of small households at a high price. These high prices are due to their location, which increases the surplus value (not to be analysed in this article). In the first quarter of 2019, one-bedroom one-bathroom apartments have increased sale prices by 7.8% in comparison to the first quarter 2018. On the other hand, three-bedroom one-bathroom houses increased their prices by 10.3% whereas three-bedroom two-bathroom houses increased by 5.5%⁷.

The household price in UF varies mainly according to its square metres, rather than the number of rooms. The most frequently advertised houses have 3 or 4 bedrooms (70%) and the most frequently advertised apartments have 2 or 3 bedrooms (35% and 43%, respectively).

Table 8 shows the descriptive statistics for every explanatory variable in the period January 2017 - August 2019.

Two interactive variables are added. First, the area interactive variable (the multiplication of district groups and m²). Second, the type of dwelling interactive variable (the multiplication of the dummy variable house/apartment and m²).

In order to create the index, we generated a linear model expressed in terms of logarithms that do not take into account the non-linearity between the square metres and the price. After calculating the monthly prices, we used the December 2018 period estimates, as stated in the Laspeyres Index methodology. This was applied as a whole including both houses and apartments and, on the other hand, separating houses and apartments so as to reach a more accurate result for each of them.

6. The imputation method applied to all the variables of the study is the imputation of conditional mean for a set of data.

7. According to Informe Trimestral de Viviendas de Portal Inmobiliario.

Graph 2 shows the volatile performance of prices in the index as a whole, including houses and apartments. This also happens in relation to the square metres, which, according to the primary source Banco Central de Chile, closed on the decline by July 2019.

By the second quarter of 2019 we observe that this trend on the decline is only related to the performance of apartment prices (see Graph 4) rather than houses (see Graph 5). This fact shows that apartments are the most important household for sale in the real estate market.

The variations in the performance in household prices with this index is similar to the previous one, for both houses and apartments. However, the structures of the two index are different. The Laspeyres methodology use the base period, December 2018, and the square metre average is maintained for estimations. The Paasche methodology, the coefficient of regression of the base period is kept, but the square metre of each period is used.

We highlight that the apartment prices (Graph 7) have a slow-paced performance in comparison to previous index, since May 2018.

Due to the fact that this index corresponds to the square root of the interaction between the previous two indicators, the Fisher methodology aims at diminishing the undervaluation and the overvaluation generated by the Laspeyres and Paasche ones in the different periods. Therefore, it is an intermediate result between both. However, as fluctuations are similar, the Fisher index does not have significant restraints, confirming the results previously mentioned.

CONCLUSIONS AND RECOMMENDATIONS

The database collected by web scraping meet the quality requirements stated by Wang & Strong (2013), e.g., credibility, objectivity and relevance. This reflects that

it is possible to build a tertiary source of information or non-traditional mechanism to collect information by contrasting sources. This article compares the studies carried out by Banco central de Chile and Cámara Chilena de la Construcción, deploying the reality of the real estate sector and explaining its dynamism through ongoing and prompt data.

Within the three methodologies, Laspeyres showed the highest, but marginal volatility. In addition, it presented a trend on the decline on household final prices in the second quarter 2019 due to both the low number of transactions and the stagnation of the construction sector, showing similar results to the ones provided by Banco Central de Chile.

For better results, the following changes could be made in the study:

- Add data to the sample collected from the website TocToc from other similar databases; for example, Portal Inmobiliario, el Rastro and Yapo. This would result in an index with a wider time span to observe more slow-paced and accurate variations in prices. Additionally, a quarterly period index could be elaborated to compare the results with the index coming from primary sources in the same period.
- Spread the index to other regions in Chile. Also, include other variables to the hedonic model of TocToc website; for instance, the amount of favourites, visits, interested people, appraised value and payment delinquency.
- Expand the analysis using rent data to generate an index that showed the characteristics increasing or decreasing the household rent price. This is based on the fact that the Ministerio de Desarrollo Social (2017) states that the amount of people renting household is on the rise.

LIST OF REFERENCES

Banco Central de Chile (2014). *Índice de Precios de Vivienda en Chile: Metodología y Resultados*. Retrieved from: <https://si2.bcentral.cl/public/pdf/estudios-economicos-estadisticos/pdf/see107.pdf>

Curiel Díaz J. (1997). La teoría de los índices de precios. *Cuadernos de Estudios Empresariales*, (7), 71-88. Madrid, España. Escuela Universitaria de Estudios Empresariales. Universidad Complutense. ISSN: 1131-6985

Centro Economía Digital CCS (2019). Tendencia del Comercio electrónico en Chile. Retrieved from: https://www.ecommerceccs.cl/wp-content/uploads/2019/04/Tendencias-Ecommerce-2019_George-Lever_eCommerce-Day-2019.pdf

Idrovo, B. y Lennon, J. (2011). Hedonic pricing models to calculate price indexes for new houses in the Santiago province. Munich Personal RePEc Archive. Retrieved from: https://mpa.ub.uni-muenchen.de/31400/1/MPRA_paper_31400.pdf

Ministerio de Desarrollo Social (2017). Resultados Vivienda y entorno. Retrieved from: http://www.desarrollosocialyfamilia.gob.cl/storage/docs/DOCUMENTO_VIVIENDA_Y_ENTORNO_RMS_CASEN_2017.pdf

Ortúzar, R. (2018). Inversiones + Planificación = Desarrollo Sustentable. *[Diapositivas de PowerPoint, 1-12, impreso]*

Parrado, E.; Cox, P. y Fuenzalida, M. (2009). Evolución de los Precios de Viviendas en Chile. *ResearchGate*, 12(1). Retrieved from: https://www.researchgate.net/profile/Eric_Parrado/publication/28321749_Evolucion_de_los_Precios_de_Viviendas_en_Chile/links/53fe76620cf283c3583bd80b/Evolucion-de-los-Precios-de-Viviendas-en-Chile.pdf

Salazar, M. y Díaz, M. (2014). Influencia del Desarrollo Urbano en el Mercado Inmobiliario de la ciudad de San-

tiago. Bachelor thesis in Business Administration, Universidad del Bio-Bio, Chillán, Chile. Retrieved from: <http://repopib.ubiobio.cl/jspui/bitstream/123456789/346/1/Salazar%20Guti%C3%A9rrez%20Mauricio%20Andr%C3%A9s.pdf>

Webster, A. (2000). *Estadística aplicada a los negocios y la economía* (3rd ed.). Bogotá, Colombia: McGraw-Hill.

TABLES AND GRAPHICS

Table 1: Results AIC for the three econometric models

Model	Obs.	ll Nul	ll Model	DF	AIC
R1	85,798	-884460	-857098.2	19	1714234
R2	85,798	-105002.8	-58599.02	19	117236
R3	85,798	-105002.8	-54762.82	19	109563.6

Source: authors' compilation.

Hint: The table show the results of AIC for three models, where the model R3 is choosed beacuse it has the least AIC.

Table2: Coefficients and explanatory variables

Lnpf	Coef.	Std. Err.
Lmt2u	.3712375***	.0080446
Bathrooms	.1994831***	.003023
Bedrooms	.0241997***	.0026261
New/old	.0581066***	.0135575
House/apartment	.1236726***	.0056413
group1	.7945399***	.0071347
group2	.4214615***	.0065206
group3	-.20647143***	.0082785
group5	-.11855838***	.0082767
group6	-.0022412	.0087934
group7	.31559121***	.0107332
area1	-.00021717***	7.82e-06
area2	-.00026471***	.0000122
area3	-.00026827***	.0000282
area5	-.00031762***	.0000234
area6	-.00026573***	.0000246
area7	-.00015332***	.0000257
C	-.0000115	.0000202
_cons	6.026263***	.0273144
Legend: * p<.05; ** p<.01; *** p<.001		

Source: authors' compilation.

Hint: The table shows the coefficients and stander deviations of real estate index, where the variables lmt2u, group1 and group2 are of the biggest contribution to the model.

Table 3: Coefficients of explanatory variables (Houses)

Lnpf	Coef.	Std. Err.
Lmt2u	.499***	.010792
bathrooms	.163***	.0035781
bedrooms	.0139***	.0035661
New/old	-.0417	.0383982
group1	.795***	.0305768
group2	.458***	.0286836
group3	-.246***	.0279983
group5	-.14***	.0283097
group6	-.0575**	.0292065
group7	.276***	.0288905
area1	-6.6e-05	.0001631
area2	8.4e-05	.0001539
area3	4.0e-05	.0001516
area5	.00013**	.0001504
area6	.00015**	.000151
area7	.00019***	.0001497
C	-.00045***	.0001499
_cons	5.7***	.0513915
Legend: * p<.05; ** p<.01; *** p<.001		

Source: authors' compilation.

Hint: The table shows the coefficients for houses, where the variables lmt2u, group1 and group2 are of the biggest contribution to the model and this have 99,99% confidence.

Table 4: Coefficients of explanatory variables (Apartments)

Lnpf	Coef.	Std. Err.
Lmt2u	.268***	.01054
Bathrooms	.273***	.0049295
bedrooms	.0189***	.0036128
New/old	.0747***	.0113992
group1	.804***	.0082358
group2	.368***	.0068867
group3	-.0509***	.0120961
group5	-.0847***	.0106857
group6	.0534***	.0100618
group7	.176***	.0815187
area1	-.00014***	8.10e-06
area2	-.00019***	.0000131
area3	-.00018***	.0000358
area5	-.00032***	.0000159
area6	-.00022***	.0000285
area7	.00015	.00002133
_cons	6.33***	.0350595
Legend: * p<.05; ** p<.01; *** p<.001		

Source: authors' compilation.

Hint: The table shows the coefficients for apartments, where the variables lmt2u, group1, group2 and bathrooms are of the biggest contribution to the model and these have 99,99% confidence. The bathroom variable has bigger contribution in apartments than houses.

Table 5: Total imputed data per variable

Variables	Amount of imputations	%
Bedrooms	1,659	2%
Bathrooms	1,865	2%
Useful m ²	24,580	27%
UF value	30,899	34%
Total	59,003	16%

Source: authors' compilation.

Hint: The table shows the imputation percentage of the variables in study, where the UF value has the highest imputation percentage with 34%.

Table N°6: Real estate advertising from 2017 to 2019

Año	Total de Publicaciones
2017	7.099
2018	34.940
2019	43.759

Source: authors' compilation.

Hint: The table shows the number of observations for each year, where the year with the most properties for sale was 2019.

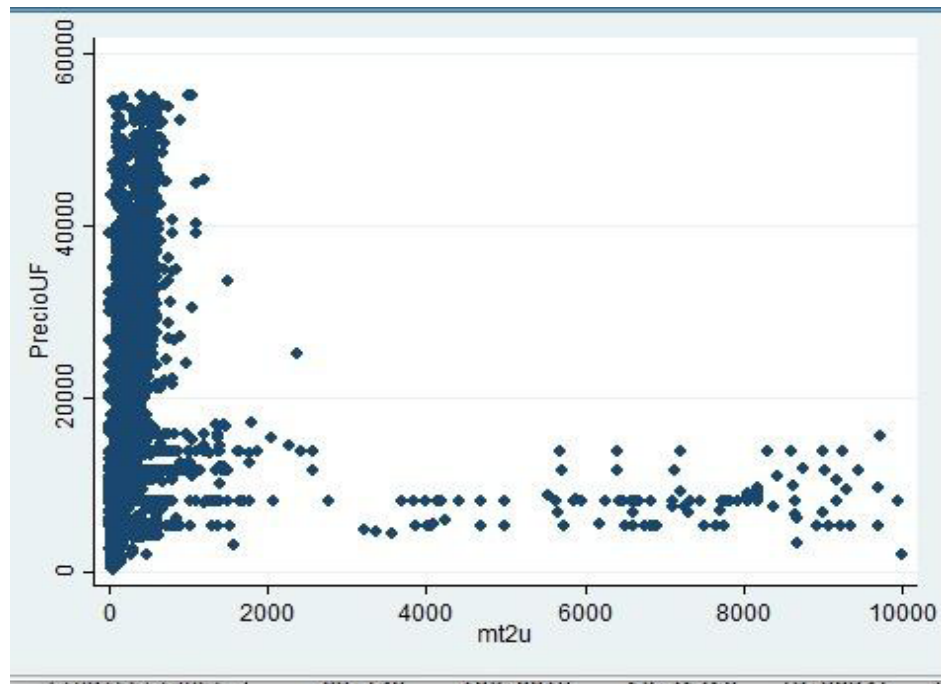
Table N°7: District groups in RM 2017- 2019

		Frecuencia		Frecuencia	
GROUP 1	Lo Barnechea	5.760	GROUP 2	La Reina	2.874
	Vitacura	5.708		Peñalolen	2.284
	Providencia	5.795		Ñuñoa	7.251
	Las Condes	13.576		Macul	1.403
	Total	30.839		Total	13.812
GROUP 3	La Florida	3.576	GROUP 4	Santiago	9.206
	La Granja	207		San Joaquín	383
	Puente Alto	3.782		San Miguel	2.008
	San Ramón	74		P. Aguirre Cerda	157
	La Cisterna	629		Lo Espejo	63
	La Pintana	145		Total	11.817
	Total	9.549	GROUP 6	Quilicura	816
GROUP 5	Cerro Navia	82		Renca	224
	Quinta Normal	682		Conchalí	271
	Lo Prado	146		Huechuraba	1.241
	Pudahuel	962		Recoleta	758
	Estación Central	1.070	Independencia	1.000	
GROUP 7	Cerrillos	312	Total	4.310	
	Maipú	3.114	GROUP 7	Otras Comunas	9.103
	Total	6.368		Total	9.103

Source: authors' compilation.

Hint: The table shows the 7 groups with their respective districts, where group 1 has the highest number observations, 30,839, continued from group 2 with 13,812. Further, the district of Las Condes, along with Santiago, are the most widely published.

Graph N°1: M² according to UF price in group 1



source: authors' compilation.

Hint: The graph shows the behavior of the Price based on mt2u, where the sample is concentrated in properties less than 2,000.

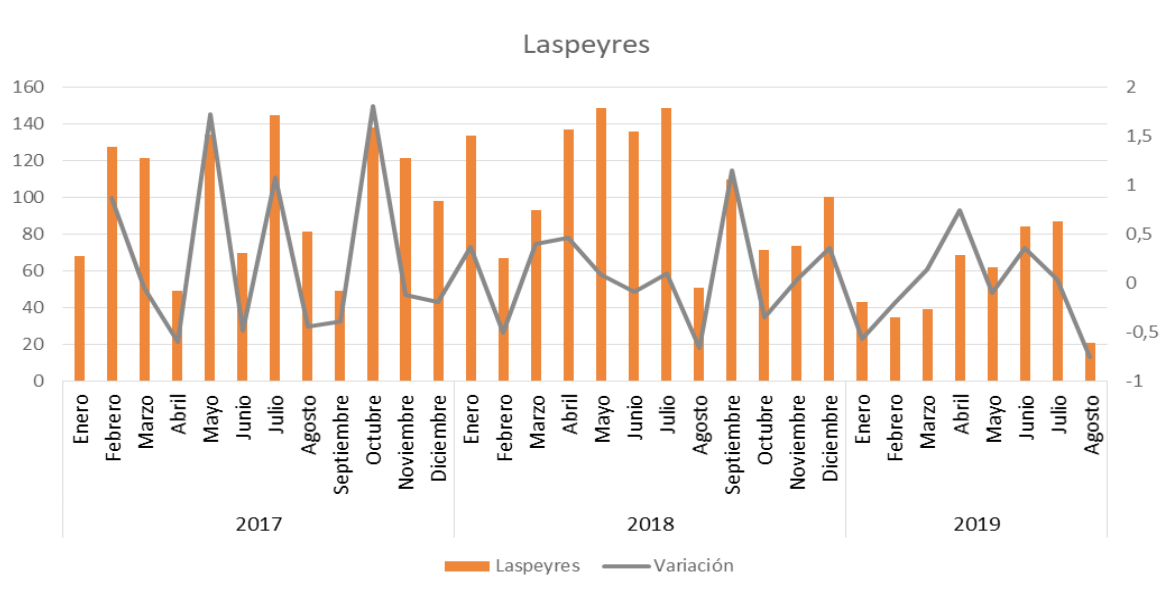
Table 8: Variables of the real estate index January 2017 - August 2019

Tipo Inmuebles - Grupos Geograficos - Variables	CASAS							DEPARTAMENTOS						
	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7
Precios UF														
Obs	11.239	6.158	7.582	1.215	4.389	2.431	8.991	19.600	7.654	1.967	10.602	1.979	1.879	112
Mean	17.366,52	9.873,38	3.902,89	5.431,24	4.141,38	5.144,65	11.065,51	9.444,52	4.826,26	3.053,99	2.962,62	2.515,98	3.263,12	6.160,39
Std.Dev.	9.405,29	5.350,35	4.535,41	5.666,02	4.915,69	4.714,42	6.946,53	4.790,34	3.330,41	3.772,25	3.231,44	2.867,19	2.909,25	6.379,27
Min	879,50	363,44	385,39	485,22	200,75	1.255,72	200,30	201,75	210,34	253,98	210,34	210,34	253,98	505,01
Max	55.000	54.885,96	54.432,83	47.190,36	54.450,31	54.415,33	54.441,62	54.573,94	54.466,23	50.000,38	54.453,93	44.017,90	44.000,34	27.207,67
M² Útiles														
Obs	11.239	6.158	7.582	1.215	4.389	2.431	8.991	19.600	7.654	1.967	10.602	1.979	1.879	112
Mean	234,56	156,944	105,33	133,21	113,22	130,84	258,01	159,50	133,51	121,24	92,43	138,42	107,72	140,3
Std.Dev.	220,07	214,86	251,90	316,60	428,59	408,39	546,95	553,25	597,72	563,25	442,83	591,67	492,08	307,44
Min	22	21	22	21	21	22	32	22	22	22	21	21	22	31
Max	10.000	8.952	9.025	9.788	9.977	9.445	10.000	9.933	9.176	8.484	8.700	6.436	8.077	3.100
Dormitorios														
Obs	11.239	6.158	7.582	1.215	4.389	2.431	8.991	19.600	7.654	1.967	10.602	1.979	1.879	112
Mean	4,19	3,91	3,35	3,49	3,34	3,48	3,80	2,69	2,41	2,54	2,01	2,23	2,33	2,29
Std.Dev.	0,93	0,92	0,86	1,11	0,87	0,95	0,95	0,81	0,77	0,64	0,79	0,75	0,76	0,74
Min	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Max	6	6	6	6	6	6	6	4	4	4	4	4	4	4
Baños														
Obs	11.239	6.158	7.582	1.215	4.389	2.431	8.991	19.600	7.654	1.967	10.602	1.979	1.879	112
Mean	3,34	2,80	2,04	1,85	1,95	2,32	3,16	2,29	1,83	1,47	1,49	1,26	1,52	1,71
Std.Dev.	1,00	0,97	0,86	0,88	0,86	1,03	1,05	0,65	0,53	0,52	0,54	0,45	0,62	0,86
Min	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Max	5	5	5	5	5	5	5	3	3	3	3	3	3	3

Source: authors' compilation.

Hint: The table shows a statistical summary about the relevant variables for the real estate index, differentiated by groups, for apartments and houses. The highest average prices UF correspond to group 1.

Graph 2: Real estate index of RM 2017-August 2019



Source: authors' compilation.

Hint: The graph shows the behavior of the generalized price index through the Laspeyres methodology. It stands out, a drop in the price level in August 2019

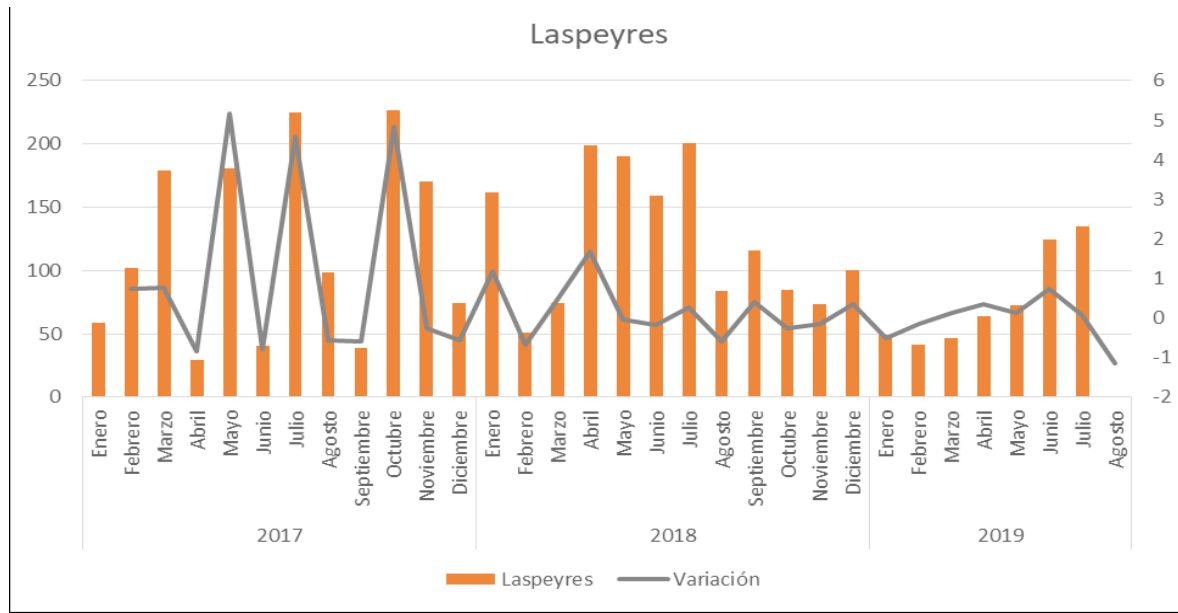
Graph 3: Household m²



Source: Banco Central de Chile, 2019.

Hint: The graph presents the variation of the surface of the houses over time.

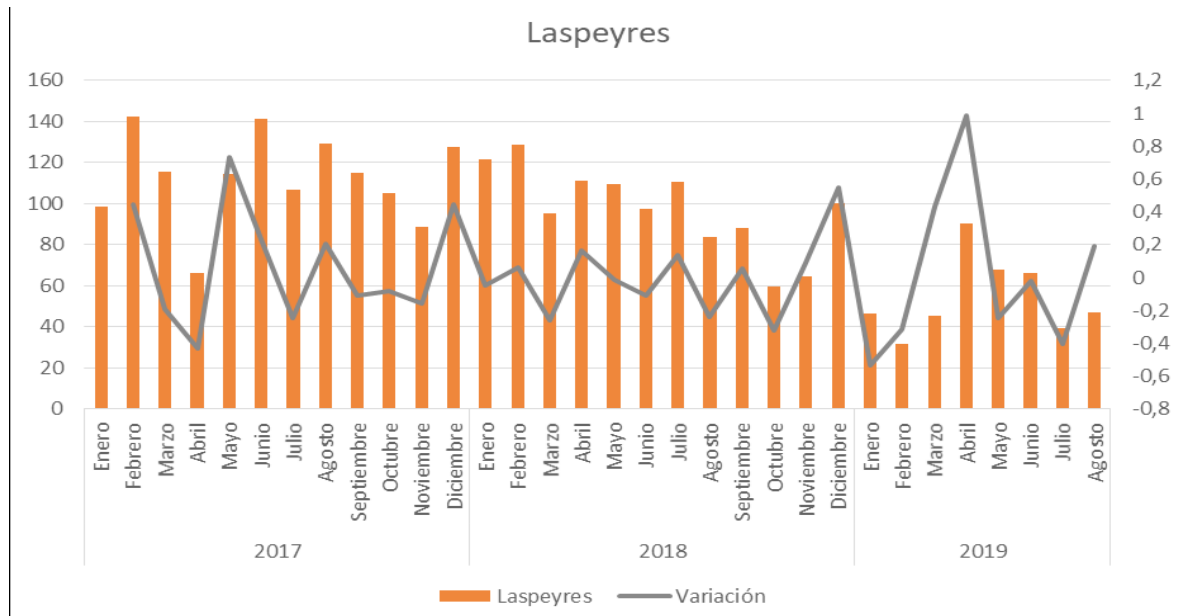
Graph 4: Real estate index of RM, between 2017 and August 2019 with Laspeyres methodology (apartments).



Source: authors' compilation.

Hint: The graph shows the behavior of the price index for apartments using the Laspeyres methodology, highlighting a drop in the price level in August 2019.

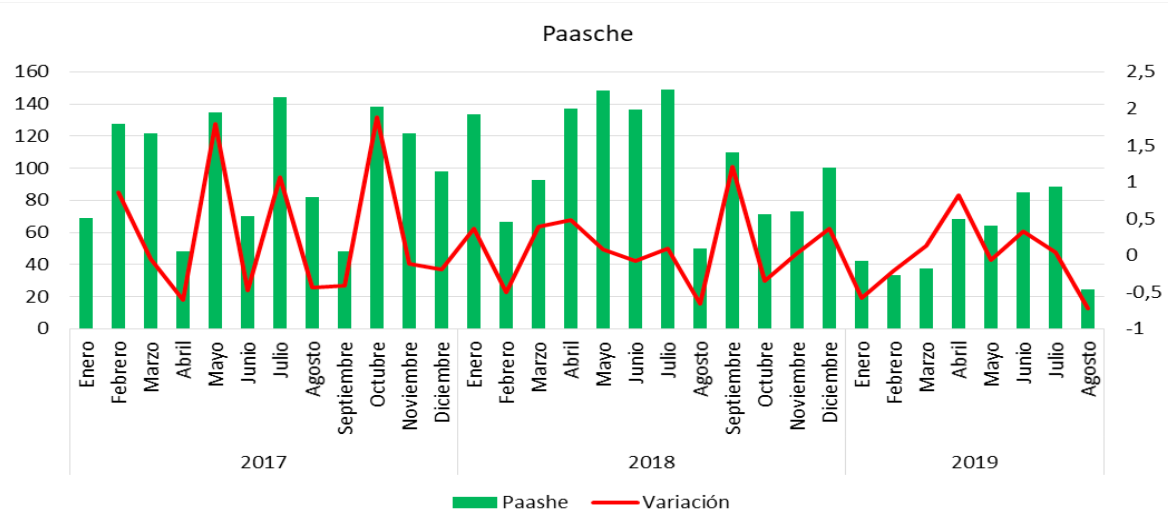
Graph 5: Real estate index of RM, between 2017 and August 2019 with Laspeyres methodology (houses).



Source: authors' compilation.

Hint: The graph shows the behavior of the price index for houses through the Laspeyres methodology, reflecting a rise in the price level in August 2019.

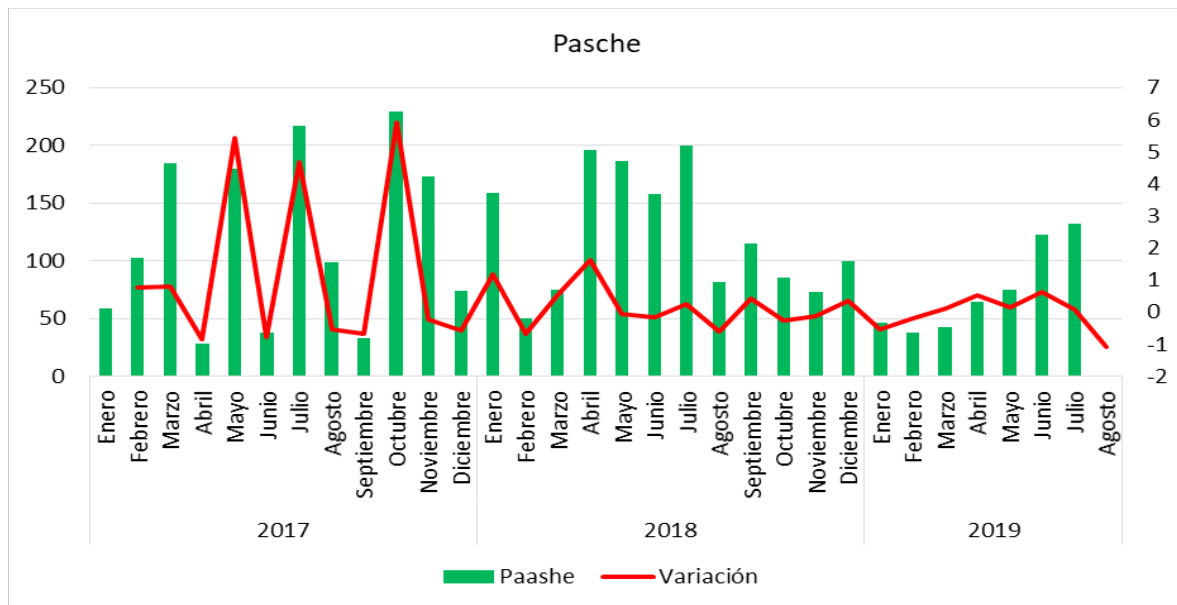
Graph 6: Real estate index of RM, between 2017 and August 2019 with Paasche methodology.



Source: authors' compilation.

Hint: The graph shows the behavior of the price index generalized through the Paasche methodology; it shows similar results to the previous ones.

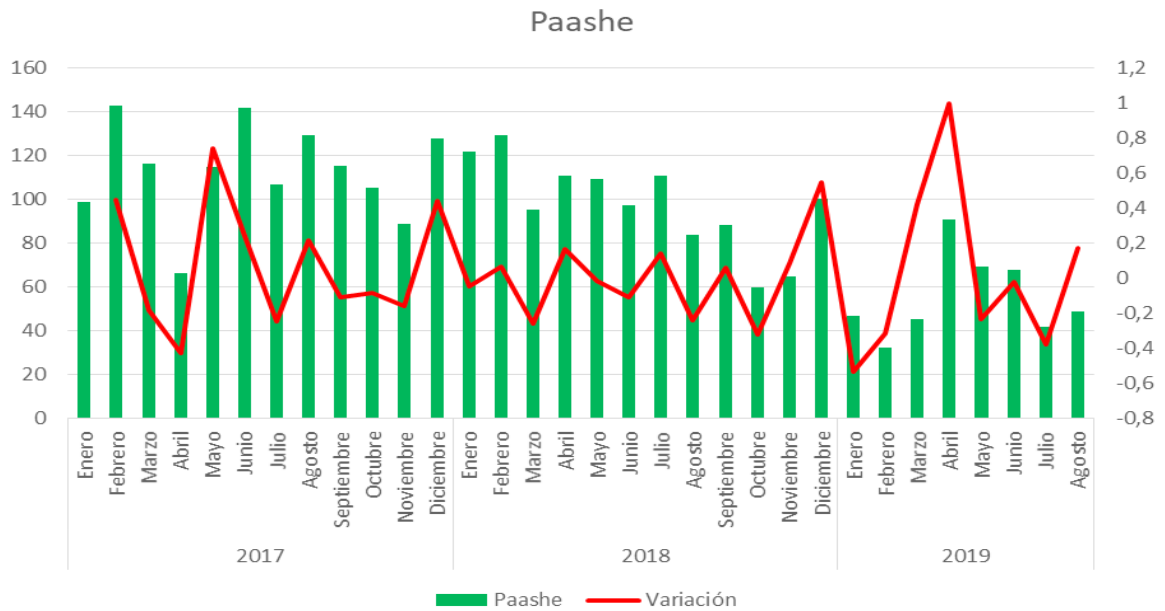
Graph 7: Real estate index of RM, between 2017 and August 2019 with Paasche methodology (apartments)



Source: authors' compilation.

Hint: The graph shows the behavior of the price index for apartments through the Paasche methodology.

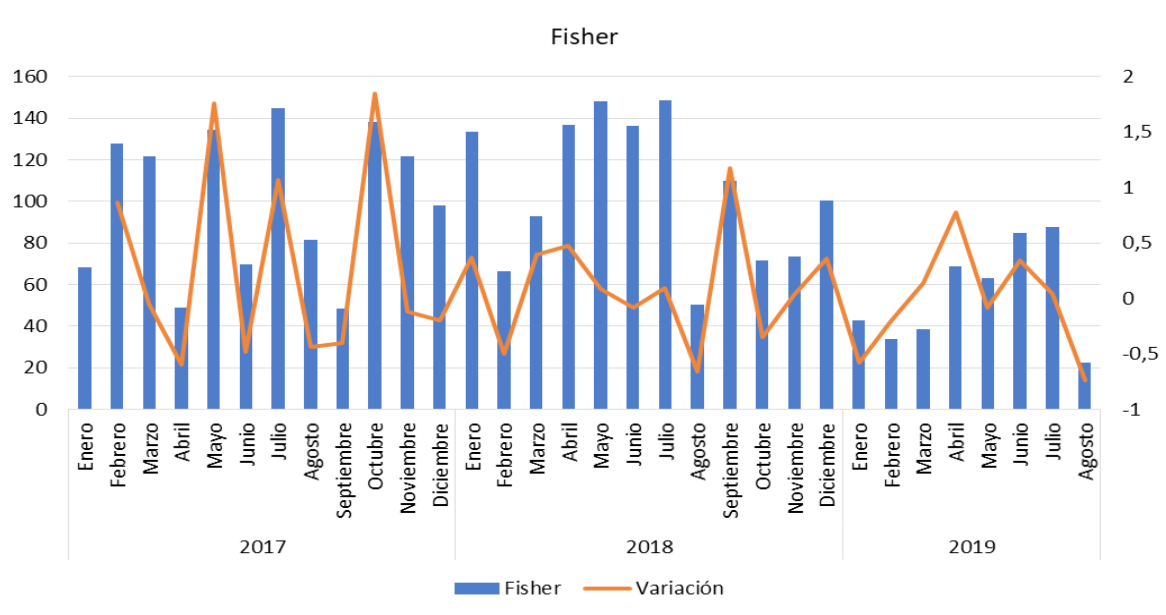
Graph 8: Real estate index of RM, between 2017 and August 2019 with Paasche methodology (houses).



Source: authors' compilation.

Hint: The graph shows the behavior of the price index for houses through the Paasche methodology.

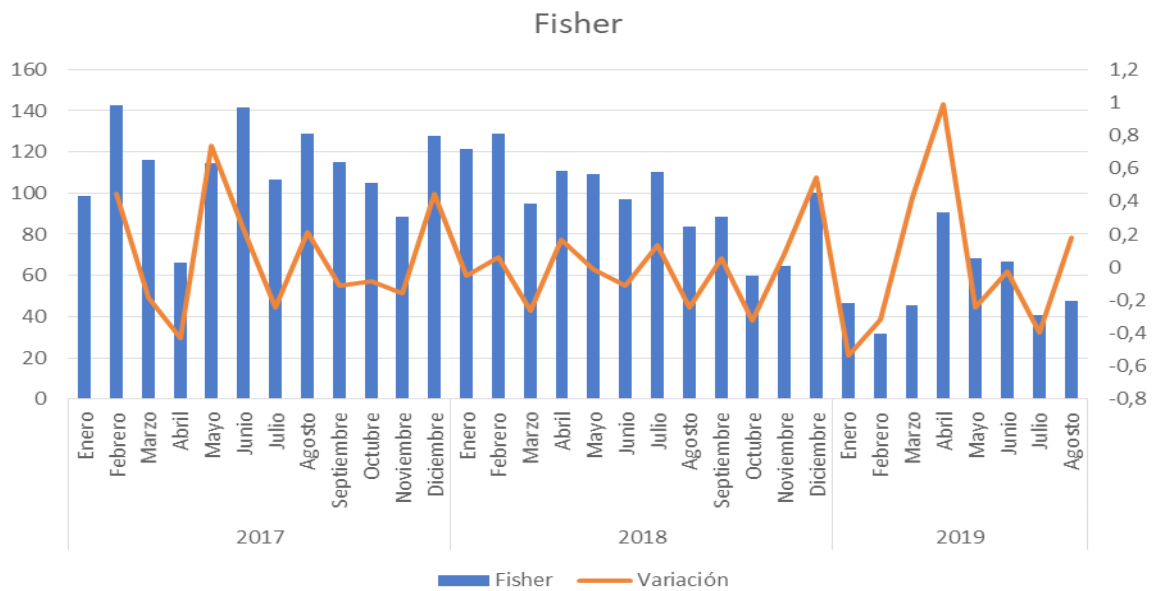
Graph 9: Real estate index of RM, between 2017 and August 2019 with Fisher methodology.



Source: authors' compilation.

Hint: The graph shows the behavior of the price index generalized through the Fisher methodology; agreeing with the results of the previous methodologies.

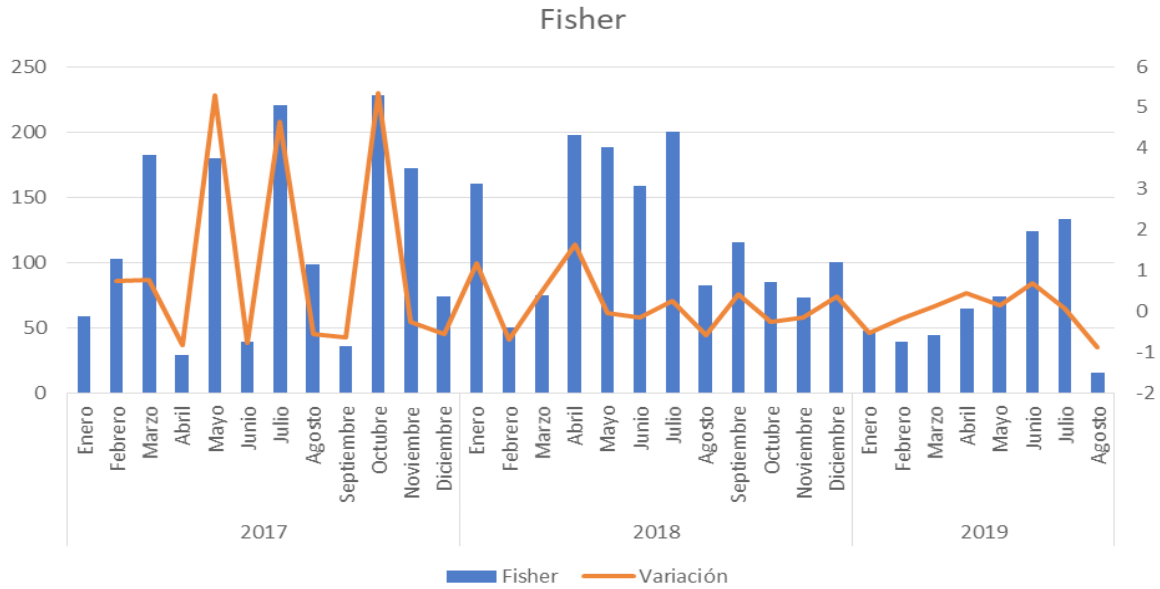
Graph 10: Real estate index of RM, between 2017 and August 2019 with Fisher methodology (apartments).



Source: authors' compilation.

Hint: The graph shows the behavior of the price index for apartments through the Fisher methodology.

Graph 11: Real estate index of RM, between 2017 and August 2019 with Fisher methodology (houses).



Source: authors' compilation.

Hint: The graph shows the behavior of the price index for houses through the Fisher methodology.